

April 2004

Creating PDF Documents

CMECF Issue Paper

US District Court, District of Utah

The Flexibility of Electronic Documents

The creation of Portable Document Format (PDF) documents is essential for the electronic filing component of CM/ECF; it is the only format that the filing system accepts. The CM/ECF project team has evaluated various options for creating PDF documents; findings and guidelines are presented below.

The CM/ECF software was designed to accept filings in PDF (Portable Document Format), in part because PDF electronically captures and preserves the formatting of a paper document. The electronic document, as displayed on screen or printed out, will look the same as the original version, regardless of the type of computer system used by the creator or viewer of the document. Page breaks, page numbering, font size, and other formatting features will be consistent, whether the PDF document is being viewed on a PC or a Mac computer, and regardless of the monitor or printer used.

Search Capability

A PDF document created in a word processing program or other application and then converted to PDF, sometimes referred to as "text PDF" document, allows the user to perform an electronic search within a document. These text documents probably encompass the vast majority of documents filed with the court. Documents that are scanned into PDF format, referred to as "imaged PDF" documents, often including exhibits or affidavits, are not searchable. The search function for text PDF documents can help judges and chambers staff quickly move to the portion of a document that interests them, simply by entering a search term.

Using PDF Document Text in Word Processing Software

A user may copy portions of text from a PDF text document into a word processing document. Among other uses, this method provides an easy short cut for creating an order based on previous filings. Instead of retyping the text, one may copy the pertinent text from a filing. The appropriate section of text can be highlighted and, using the software's various cut and paste tools (some experimentation may be needed to choose the tool that best preserves the formatting), the selected text can be lifted out of the PDF document into your word processing document. Imaged PDF documents will not permit copying of text into other applications.

Background

There are two primary methods for creating PDF documents: formatting text documents into PDF at the time of creation or scanning imaged documents from paper into PDF. The former method offers at least four significant advantages over the latter.

Reduced Labor: Imaging is a labor-intensive task and if documents must be imaged, either the

attorney or the court will have to provide that labor. Courts should strive to limit the amount of imaging done by court staff.

Improved System Performance: Text documents are much smaller than imaged documents. For a typical document, its PDF text version would be only 20% of the size of its imaged version. Therefore, whenever an imaged document is stored or viewed, it puts five times more load on the network than it would as a text document. It also take five times as long to transfer. That extra transfer time is particularly noticeable when the user is working over the DCN or the Internet.

Reduced System Storage Cost: An imaged document consumes about five times more storage space than it would as a text file.

Text Search Capability: Text documents can be searched for words and phrases (as is done in Westlaw or Lexis). Imaged documents cannot be searched without first performing an extra step called optical character recognition (OCR). OCR is labor-intensive and does not yield 100% accuracy.

Text Copy Capability: Text may be copied from a text PDF document and pasted into an other application, such as word processing software.

Creating a fully electronic case file is a worthy goal, and to do that some imaging is necessary. However, it might not be possible to convert some exhibits and other pre-existing documents into an electronic format. For other documents, conversion into electronic format might be possible but not economic. For example, if a large document is not likely to be viewed, or will be very seldom viewed, there may be little benefit to offset the cost of scanning it. The Court's Administrative Procedures will provide that some documents may be filed in paper format.

Creating PDF documents from a word processing package

The optimal method for creating a PDF document for filing in CM/ECF is a simple method: create it directly from a word processing application using Adobe Acrobat's PDFWriter, PDF Distiller, pdfFactory, or another product capable of producing pdf documents. PDF files created in this way have two advantages: they are much smaller in size than documents that have been scanned and they are text-searchable and allow text to be copied into other applications. The District of Utah recommends this method for creating a PDF file from a word processing application.

Adobe Acrobat

The price of Adobe Acrobat is approximately \$300. The latest version, 6.0, has been tested by the CM/ECF project team, and it is recommended that the compatibility be set to version 4.0 or 5.0, to ensure full compatibility. This is because documents created with version 6.0 can be accessed by previous versions of Adobe Acrobat Reader, but may give the user a prompt to upgrade to a more current version. The Reader is a free download (but allows only the ability to read, and not create pdf documents.)

The national CM/ECF project team also tested the integrity of 32 fonts when converted from either WordPerfect version 8 or Word 97 into PDF using Adobe Acrobat 5.0 PDFWriter. The tests showed that while characteristics of some fonts were lost in the conversion, every font conversion produced legible results. Firms will want to ensure that the fonts that they normally

use will produce adequate and legible results upon conversion.

Other PDF Creation Software

Other manufacturers provide alternative software for creation of PDF files. Some of these which have shown to be an acceptable method of producing pdf documents at a much lower price than Adobe Acrobat are shown below

Brand name	Price	Internet URL
PdfFactory	\$49.95	http://www.fineprint.com/
CutePDF	Free	http://www.cutepdf.com/
PDF 995	\$9.95	http://www.pdf995.com

More information on PDF creation software is available at
<http://www.pcmag.com/article2/0,4149,1195058,00.asp>

WordPerfect

WordPerfect versions 9 and above include a feature (“Publish-to-PDF”) that allows a document to be converted directly into PDF format. This conversion tool is an acceptable method of PDF creation for the CM/ECF system.¹

Microsoft Word

Microsoft Word 97, Word 2000, and Word 2002 do not currently provide a PDF converter and the Product Guide for Microsoft Office XP makes no reference to it. There is, however, a macro installed with Adobe Acrobat Standard and Professional (beginning with the 3.01 version) called PDFMaker that provides enhanced features for creating PDF files from Word. It is installed by Acrobat in the Microsoft Office/Office/Startup folder and is accessed through the File/Create Adobe PDF menu item or through an Adobe icon on the toolbar. It uses Acrobat

¹ Testing by the national CM/ECF project team shows that an eight page document converted to PDF with WordPerfect 9.0’s Publish-to-PDF feature is two times larger in size than the same document converted with Acrobat PDFWriter. When a one page document with an image embedded in it (a small agency seal) is converted, the difference in size is much greater; the PDF file created by WordPerfect is nearly 100 times larger than the file created with PDFWriter. This difference in size can be lessened by changing the bitmap compression setting from the default of ZIP to JPEG (through the Details tab) and the quality factor from 2 (the highest quality) to 255 (the lowest quality). However, even with the quality set to 255, the file is still three times greater and with the quality set to 2, the file is 11 times greater. These tests were done with the Advanced Output setting set to the default of RGB; changing this setting to gray does not significantly change the size of the file. The quality of the seal is better at the highest quality setting (and better than the quality of the seal created through PDFWriter). Preliminary testing of this feature in WordPerfect 10 shows that improvements have been made in the size of the file created. However, compared to the file created by PDFWriter, the WordPerfect-created PDF file is still nearly three times greater for a plain text file and eight times greater for a file with an image embedded in it. The size of the imaged file can be diminished to about four times greater by changing settings.

PDFWriter or Distiller and converts Word features such as headings to PDF bookmarks, URLs to PDF Weblinks, crossreferences within a document to PDF links, etc. A full list of these features can be found at the www.adobe.com web site (search for PDFMaker) or for users that have already installed

Adobe Acrobat, a help file can be found in the Program Files/Adobe/Acrobatx.0/Help/ENU folder. (Adobe Acrobat 4.0 users who wish to use PDFMaker with Word 2000 must upgrade [this is a free upgrade from Adobe] to Acrobat 4.05; the PDFMaker that shipped with Acrobat 4.0 does not work with Word 2000.) PDFMaker is not available for the Macintosh operating system.

The PDF files created in Word using the File/Print to Acrobat PDFWriter method and the File/Create Adobe PDF (the PDFMaker macro) method are nearly identical in size. There do not appear to be any differences in the quality of the PDF files produced, but it has not been fully tested by the CM/ECF project team. The advantage of the PDFMaker macro is the additional conversion features that it provides for Word documents. If these particular Word features are not used, the File/Print to Acrobat PDFWriter method is adequate.

Creating PDF documents from a scanning system

For those documents that must be imaged because a word-processed version does not exist, the preferred method is to scan the document directly into PDF format using Adobe Acrobat (versions 4.0 and above provide this feature). (From the File menu, choose Import/Scan.) Scanning to a format other than PDF (e.g., TIFF) would add both delay and labor, requiring not only the scan but then a conversion to PDF.

Factors to consider when scanning

There are several factors to consider when determining how the scanning will be done.

- The quality of the document when scanned is determined by the level of detail recorded by the scanner. This detail is referred to as the resolution, which is measured by the dots per inch (dpi). 200 dpi is the recommended resolution. Less than 200 dpi should not be used.

A higher resolution:

- is slower to scan (this is dependent also on the scanner and the number of pages that are scanned).
- creates a document with a larger file size. This, in turn, causes the loading of the document into the case management application and the retrieval of the document from the application, to be slower.

- Another factor to consider is the mode of scanning. Almost always, it should be done in black and white, so that the file size will be as small as possible. However, there are some documents (those with shaded boxes, for example) that may need grayscale instead; this, however, will produce a file that is much larger in size than the document scanned in black and white. Never scan in color, unless absolutely necessary (perhaps for exhibits originally produced in color). This, too, will produce a document with an extremely large file size, which will take longer to load into the application and longer to retrieve for viewing purposes.

Comparison of text vs. scanning file sizes

A 50 page text-only document was converted to PDF in text format and scanned format, using several different packages. In this table are the results.

50 page text document converted	Resulting size in KB
Adobe Acrobat 5.0	94
Cute PDF (free download)	156
WordPerfect	179
pdfFactory	151
50 page scanned document	
Paperport to tif format	2,172
Paperport to pdf format	1,257

Limits on size

Because of the staff time, disk space, and long retrieval time consumed by large documents, the District of Utah has set limits on the size of documents and number of pages that will be accepted electronically and the file size that will be accepted. Documents should not exceed 2 MB in size. A scanned document should not exceed 40 pages in length (approximately 1.5 - 2 MB in size). Longer documents should be broken up into logical sizes of 40 pages or less, and filed electronically as attachments. Text documents created from a word processing program should also not exceed 2 MB in size. These could range in size from 40 to approximately 150 pages in size, depending on the amount and type of text in the document.

Summary of Differences - scanned pdf vs. text document converted to pdf

	Text Document Converted to PDF	Document Scanned to PDF
File size	Small - better	Large - worse
Key word searchable	Yes	No
Copy text and Paste into another document	Yes	No
Clarity of image	Printer quality	Fax quality